Putting It Together Online: Information Need Identification for the Domain Novice User

Charles Cole, John E. Leide, Andrew Large, and Jamshid Beheshti
Graduate School of Library and Information Studies, McGill University, 3459 McTavish Street, Montreal, Quebec, Canada H3A 1Y1. E-mail: {charles.cole, john.leide, andrew.large, jamshid.beheshti}@mcgill.ca

Martin Brooks
Institute for Information Technology, National Research Council of Canada, Montreal Road, Building M-50, Ottawa, Ontario, Canada K1A 0R6. E-mail: martin.brooks@nrc.ca

Domain novice users in the beginning stages of researching a topic find themselves searching for information via information retrieval (IR) systems before they have identified their information need. Pre-Internet access technologies adapted by current IR systems poorly serve these domain novice users, whose behavior might be characterized as rudderless and without a compass. In this article we describe a conceptual design for an information retrieval system that incorporates standard information need identification classification and subject cataloging schemes, called the INIIReye System, and a study that tests the efficacy of the innovative part of the INIIReye System, called the Associative Index. The Associative Index helps the user put together his or her associative thoughts—Vannevar Bush's idea of associative indexing for his Memex machine that he never actually described. For the first time, data from the study reported here quantitatively supports the theoretical notion that the information seeker's information need is identified through transformation of his/her knowledge structure (i.e., the seeker's cognitive map or perspective on the task for which information is being sought).

Introduction

The interaction model used by Web search engines today assumes that the user's information need is static during a user search session (Hearst, 1999), that the user's query to the IR system is simply a one-to-one translation of the need into conceptual terms, and that the information need enables the user to make accurate relevance judgments when looking over the IR system's output in the results list. However, research has shown that in many information situations neither the query formulation assumption (Savage-Knepshield & Belkin, 1999) nor the relevance judgment assumption (Saracevic, Mokros, & Su, 1990) is true. This is because the primary postulation of the presence of an identified information need when the individual begins information seeking is itself often untrue.

The domain novice exploring a topic to gain a personal understanding of it (Kuhlthau, 1999) is an example of a case where the user seeks information when the information need is not yet identified. In these cases, information need is present at an unconscious level (Taylor, 1968) as an anomalous state of knowledge (ASK) (Belkin, 1980; Belkin, Oddy, & Brooks, 1982), or as a perceived gap in understanding (Dervin, 1992). The gap involves what Kuhlthau (1993) has called "cognitive uncertainty." Cognitive uncertainty causes, or is immediately followed by, affective uncertainty, which leads to the commencement of information seeking to relieve the affective uncertainty or anxiety. In the exploration stage of seeking information—in Stage 3 of Kuhlthau's six-stage Information Search Process (ISP) model—the user's uncertainty actually increases (Kuhlthau, 1993). And yet it is at this point where it often is necessary for the domain novice user to go to an IR system to begin seeking and finding sources of information.

Information Need Identification

Information need is identified via interaction with information objects (Savage-Knepshield & Belkin, 1999). In the pre-Internet era, when the information accessing system (i.e., the library's catalog) and the information itself were physically separated, identification of information need occurred via interactions with information objects the user read and thought about offline (Bates, 1989). In the Internet environment, however, where more and more information sources are instantaneously available online by clicking on hypertext, information need identification that is concurrent

Received January 9, 2003; revised September 17, 2003; accepted May 13, 2004

 • Published online 16 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20156

with the online accessing of information will play an increasingly larger role in improving IR system performance.

The general problem we present here is how to design an interactive IR system for a domain novice user of an IR system, who is accessing the system before information need identification has taken place. User interaction with an IR system at a pre-need identification moment can be extremely frustrating; the domain novice user frequently freezes up, unable to change tactics or strategy when confronted with an access barrier while online using the system (Brajnik, Mizzaro, Tasso, & Venuti, 2002). This user cannot formulate an appropriate query to the system, and cannot distinguish good from bad in the results list.

The domain expert, on the other hand, has an internal compass from his/her home domain, as well as knowledge about the structure of knowledge, the structure of the information store, and about searching using IR systems (Hjorland, 1997, p. 167; cf. also, Pennanen & Vakkari, 2003; Robertson, 2001). An information need identification IR system must somehow communicate these structures to the domain novice user in a way that they can be understood and assimilated.

The need identification in such a system, we predict, will be produced by the domain novice user interacting online with multiple collocation and collocation-like accessing technologies; these accessing technologies can be considered in themselves information objects, which we define as objects that cause need-focusing information processes to occur in the user.

Our purpose here is to present a conceptualization of an information need identification system, called the INIIReye System, that enables users to concurrently identify their information need while they are accessing information. The design of the INIIReye System is based on library information need identification systems codified in subject cataloging and classification rules in the last half of the 19th century, as well as current Internet information need identification technology that includes both adaptations of library accessing systems and hyperlinked text. Internet hyperlinking is difficult to index as it creates a unique path through Internet sites and pages, which is chosen by the user in an ad hoc manner. Here, we present a design for such an index, called the Associative Index. In a first step to testing the efficacy of the design of the Associative Index, we report the results of a study of 59 undergraduates in a naturalistic information need setting who are seeking information for their course essay.

Library Antecedents

In the pre-Internet, library environment, the trend towards open stacks and self-service in the late 19th century (Cutter, 1891–93, p. 5) caused catalogers, indexers, and classificationists to become more concerned with assisting the user to (a) find, (b) identify, and (c) select needed information while using the library's IR system (the card catalog). The three objects (objectives) were expressly codified in 1876 by Cutter (1876/1904) in the introduction to his *Cataloging Rules for a Dictionary Catalog*:

1. To enable a person to find a book of which the author, title, or subject is known
2. To show what a library has by a given author, on a given subject, in a given kind of literature
3. To assist in the choice of a book, as to its edition (bibliographically), or as to its character (literary or topical)

Since their publication, the first object has been referred to as the finding object for known-item searches, the second object has been referred to as the collocation or identification object for unknown-item searching, and the third object has been referred to as the selection object (Svenonius, 2000, pp. 15–16). A fourth object, the obtaining object, was added in 1997 by the International Federation of Library Associations and Institutions (IFLA); and for the interactive technology era, Svenonius (2000) has recently added a fifth navigation object. We intend to analyze each object in a separate article. We discuss the fifth object in Leide, Large, Beheshti, Brooks, and Cole (2003). The second object, the collocation or identification object concerns us here.

The Latin root of the word "collocation" is "Placing things together" (Latin root: "col-(con-) together + locare to place"; Oxford English Dictionary, 1989, p. 487). Cutter's second collocation object was meant to facilitate user identification of needed information by collocating or bringing together similar items by author, type of literature, and subject. A second meaning of collocation is "arrangement." Showing the user an arrangement scheme in the catalog, to indicate relationships between items, has always been difficult (Carlyle, 1997, pp. 88, 92). In an alphabetical catalog, from the user's point of view, arrangement is primarily accomplished by the library's classification scheme (Taylor, 2000, p. 267).

Classification schemes facilitate user identification of their information need by placing books together on the shelf according to "a hierarchy of genus and species, class and subclass, that progresses downward from general to specific" (Shera, 1965, p. 78). The assumption at the root of this accessing technology is that in the human intellectual process things "have likeness or unity and by this likeness or unity are set in relation to one another" (Sayers, 1959, p. 79). As a result, "When [the library user] looks among the shelves of the library, he should find there what he was only vaguely conscious of wanting; indeed, it is only then that he will be able to realize exactly what it is he wants" (Ranganathan, 1962, p. 17). The user, for example, will ask to see books on birds rather than books on blue jays, the actual information need (Donovan, 1991; cf. also, Bates, 1986, pp. 365, 366); but because of the classification scheme, the books on blue jays are placed near the general bird books, thus allowing the user's eye to fall upon them, and to recognize his or her information need from their advantageous placement.

The Internet: Bringing, Placing, and Putting Things Together

In the current Internet environment, search sites use three types of collocation-like accessing technologies: (a) the subject directory, (b) the search engine search box in the results list, and (c) hyperlinked text technology. The first two types are adaptations of library-era classification and subject cataloging accessing systems, respectively. We examine hyperlinked text in some detail below as it is the accessing technology at issue in the study reported in this article. We will examine subject directories and search engines in detail at a later time, only briefly commenting on them here.

Subject directories. With some exceptions (e.g., KidsClick.org), most Internet subject directories are not pure classification schemes but rather alphabetico-classed arrangement schemes (Weinberg, 2002), so they do not aid the domain novice's identification of information need by bringing related subject disciplines to this user's attention (University of Maryland, 1998) (i.e., subject directories are intracategory rather than intercategory focused; e.g., Google.com, 2003). We retain the interrelating feature in our conceptualization of an information need identification IR system (shown below).

Search engine box. The search engine's particular algorithm collocates or clusters system output in the search results list (Meadow, Boyce, & Kraft, 2000), but the problem for a domain novice user in subject-arranged Internet results lists (e.g., Northernlights.com or AltaVista.com's AltaVista Prisma) is that such results lists are retrieved based on links between sites, and user frequency of such links. Because these links are designed to change over time, they do not show the user the basic structure of the subject discipline that is constant over time. They are thus difficult for the domain novice to learn and use for information need identification purposes.

Indexing Hyperlinked Text: The Problem

Hyperlinked text is a third accessing technology on the Internet. The hyperlinking effect of mixing information retrieval with the concurrent perusal of and interaction with information contained in the source document is a truly innovative information retrieval feature found in Internet searching. However, a hyperlink-based information access is notoriously difficult to index (Hert, Jacob, & Dawson, 2000, pp. 972, 973, 981; cf. also, Tebbutt, 1999).

Vannevar Bush (1945), in a description of his Memex machine, a precursor to the Internet (Nyce & Kahn, 1989), spoke directly to this indexing problem. He envisaged a new profession, a mixture of librarian and researcher, called "trail blazers," who would form fixed sequential trails through an information topic for others to follow (for the HYPERCATalog, a hypertext-based system with fixed trails, cf., Bjorklund, Olander, & Smith, 1989). The fixed sequence vision of information seeking, however, was subsequently rejected by the early developers of Internet hyperlinking like Nelson (1972/1991, p. 253), who said "no sequence need be imposed on the material. . . . Such non-sequential or complex structure we may call hypertext."

Nonsequential hypertext is at the heart of present day Internet information seeking, and Bush's idea of a "trail blazing" profession has been dropped. The trail-blazer concept in Bush is actually a secondary issue. The main purpose of the Memex was: "to extend and support the power of human memory and association" (Nyce & Kahn, 1989, p. 216). The human mind works by association: "With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain" (Bush, 1945, p. 106). Hyperlinking gives "instant and potential" information support to the user's associative thinking by providing immediate access to the information store when a textual element in the Web page the user is reading stimulates associated thoughts. However, the hyperlinks followed by the user do not form an index but rather a navigational structure or a record of the user's search session. Indexing the user's associative thinking, a cognitive activity that is only partially expressed physically by the user's action of clicking a hypertext and hyperlinking to an informational source, is extremely difficult.

Indexing Hyperlinked Text: A Solution

Hert, Jacob, and Dawson (2000, p. 981) suggest that indexing hyperlinking would have to be done through the "introduction of intermediate indexing structures." Such an intermediate structure would serve as a go-between structure linking the user's associative thinking with system accessing schemes based on collocation technology. We set about designing such an intermediate structure, we call the Associative Index, starting from Bush's partial description of associative indexing: "... the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the Memex. The process of tying two items together is the important thing" (Bush, 1945, p. 107).

As just described, the two fundamental notions of Bush's partially described associative index are (a) the notion of selection, and (b) the notion of two items being tied together in the person's mind for selection to occur.

Bush is silent on the issue of how the selection is made or any further description of an associative index.

Description of Associative Index

The Associative Index is user- and task-specific; in our case it is intended for history undergraduates exploring their essay topic before they have identified their information need. The basis of the Associative Index is provided by all

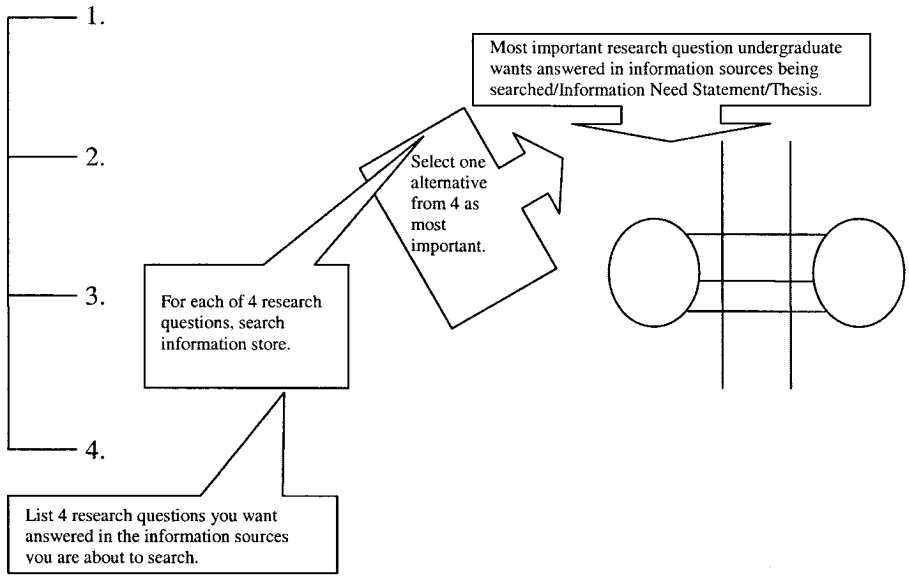

FIG. 1. Associative Index. The tacit knowledge collector is shown on the left-hand side of the diagram; the Index is shown on the right.

the associated bits of tacit knowledge the undergraduate associates with the topic of the essay. In Figure 1, these are obtained by asking him/her to list four research questions he/she wishes answered by the information source (shown on the left-hand side of Figure 1), to interact with the information source for each research question, to select the most important research question, and finally to collocate or bring together all the essential subject terms from the selection in an index structure or schema that we call here a constructed information need perspective index schema (shown on the right-hand side of Figure 1). The index then serves as a query formulation device, from which the undergraduate can carry out the final information search before leaving the system. The assemblage of data bits occurs as shown in Figure 1 (left-hand side of the figure), while the indexing part occurs as shown on the right-hand side of Figure 1 (two circles joined by three horizontal lines).

The Information Need Identification System

The Associative Index just described fits into a total information need identification system for the domain novice user. In effect, the Associative Index "puts-together," brings together, or collocates the tacit knowledge the domain novice already knows about his/her essay topic, to index it. But information need is identified by the user's own tacit knowledge about the subject being searched and interaction with the knowledge store (Bates, 1989; Hjorland, 1997). To provide a global context for a design solution to the General Problem of the domain-novice user using an IR system before information need identification has taken place, the Associative Index is placed in a total scheme with three other collocation-type index schemes, two of which are domain or objective representations of the topic area being searched. The Associative Index serves as a user-controlled executive mechanism between the user and these two domain representations of the topic area. Together, the four collocation-type schemes fit together into an information need identification system, called the Information Need Identification Information Retrieval eye System (INIIReye System—pronounced "inner eye system").

The INIIReye System presents "many IR technologies to the user, to promote user cognition" (Hendry & Harper, 1997). It relies on stricter formulations of library accessing technologies than is currently the case on the Internet, with formal, controlled vocabulary subject terms (i.e., syndetic maps, with narrower and broader terms), and a classification system that indicates the subject domain's structure to the user (cf. also, Bates, 1986; Bjorklund et al., 1989; Brajnik et al., 2002; Carlyle, 1997). The four schemes are summarized below and depicted in Figure 2:

- The Associative Index, which acts as an intermediate structure between user and the other index/classification schemes in the INIIReye System, brings all the user's associated thoughts together. This index serves as a user chart which links the user to system schemes for the purpose of creating prior conditions for information need identification. The index also contains a selection component (Figure 2d). The Associative Index changes form from the beginning (Figure 2a) to end of the search session (Figure 2f), as the user gradually identifies his or her information need.
- The Classification Scheme places or arranges like items together for user browsing (Figure 2b).
- The Subject Cataloging Collocation Scheme, which brings like items together and then summarizes and visualizes the arrangement as a syndetic map (Figure 2c). For a description of this scheme, cf. Leide et al. (2003).
- The Navigation Scheme links together the user's navigation history for one search session, providing a navigational chart (Figure 2e).

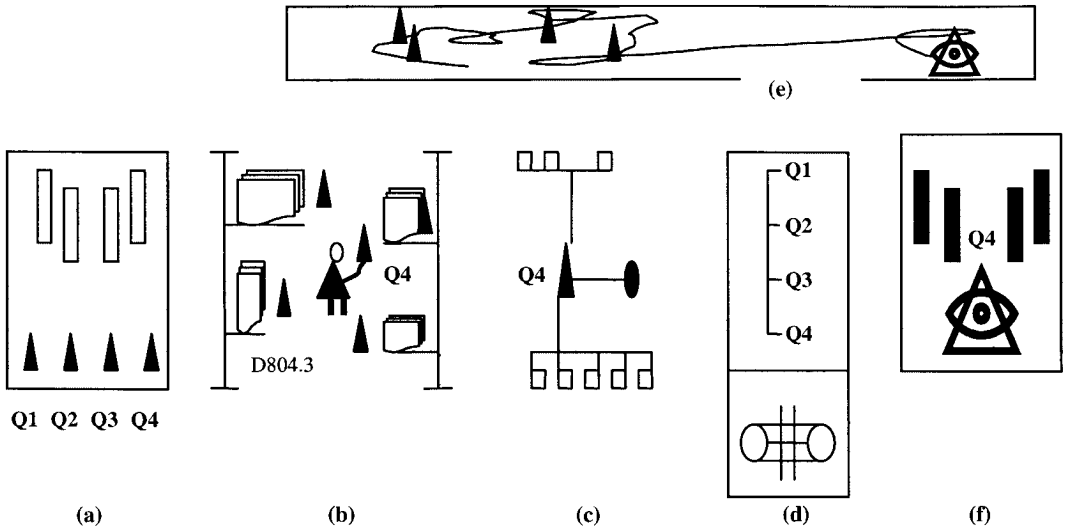

FIG. 2. The INIIReye System for a domain novice user in the early stages of his or her ISP. (a) Associative Index (user chart). (b) Classification scheme. (c) Subject cataloging collocation scheme. (d) Associative Index (tacit knowledge collector, top; index, bottom). (e) Navigation scheme (navigation chart). (f) Instantiated Associative Index, providing the user with an identified information need schema for seeing in his or her mind's eye INIIReye System output as meaningful (or not) to the task at hand.

The classification and subject cataloging collocation schemes are system determined, while the Associative Index and the navigation scheme are user determined. The Associative Index acts as an intermediary indexing structure between the user and the two system indexes. All four schemes essentially bring like items together in different ways, but the Associative Index also creates a cognitive vehicle for the user to put together his/her thinking, thus creating the prior conditions necessary for information need identification and successful information searching to occur (cf. also, Ellis, 1989; Drabenstott, 2003).

Specific Problem

The specific problem addressed herein is to test various hypotheses, formulated here as research questions, for the design of an Associative Index. Only the "tacit knowledge collector" part of the Associative Index is addressed in the following study (the left-hand side of Figure 1). The Associative Index is an "intermediate indexing structure" for a hyperlinked Internet information retrieval environment, based on Bush's missing associative indexing concept. The Associative Index is described above as part of the INIIReye System, and shown in Figure 1, and in Figure 2a, d, and f, but it is also a system itself, which, rather than a single listing of index terms, takes the user through a series of cognitive activities that we hypothesize promote information need identification.

The Study

In the winter term of January to April 2003, 59 undergraduate students attending four separate history and psychology courses at McGill University were interviewed using the interview schedule shown in The Appendix. McGill University is a large (26,000 fulltime students), North American research university offering doctoral programs in a full range of disciplines, including medicine and law. The 59 students were interviewed about their topic for a course essay, an essay that would be evaluated by the course instructor and used as a component of the total mark the student would receive for that course on the student's transcript. The interviews, therefore, were about the student's real information need.

The students in the four courses represented both the humanities or "soft" end of the social science spectrum (history students) and the science or "hard" end of the social science spectrum (psychology students) (cf. Whitmire, 2002), as well as both inexperienced first-year students and more experienced third-year students. Twenty students were from one large (over 150 students) 200-level (1st year) history course, 20 students were from one large (over 100 students) 400-level (3rd–4th year) psychology course, and 9 and 10 students, respectively, were from two small (10–15 students) 400-level (3rd–4th year) history courses. We will complete a 2×2 cell structure study design by adding a first-year psychology course with an essay requirement in the fall term, 2003 (such a course was not given at McGill during the term when the study reported here was conducted).

The researchers entered the four courses in the first or second week of classes and gathered volunteers for the research project. The sample of students was self-selected and thus does not necessarily represent the larger population of history or psychology undergraduates at McGill University or elsewhere. We then contacted the volunteers by e-mail, asking them to contact us when they had selected their essay topic and had done some preliminary thinking about it (i.e., when they were in Stage 3 of Kuhlthau's, 1991, Information Search Process [ISP] Model). The interview schedule required the students to list subject terms or to draw visualizations of their topic on large sheets of newsprint-type paper. The interviews were video/audio-taped and lasted from 30 minutes to 1 hour.

Research Instrument

In the study, some of the cognitive activities in the Associative Index are encapsulated in an interview schedule, shown in The Appendix. The research intervention, the interview, occurred just after the subject had selected his or her essay topic. The interview schedule's overall purpose is to facilitate the undergraduate's narrowing of his or her initial Topic Statement (Question 1 in the interview schedule) into an Information Need Statement (Question 9). We make the assumption that the subject gives an Information Need Statement at the end of the interview. In this study, we gather exploratory evidence in support of this assumption. In the interview, we do not label the last statement in the interview given by the subject concerning "what your paper will be about" as their Information Need Statement because this is a technical information science term that would have confused the research subjects.

The interview schedule facilitates the subject's narrowing of his or her Topic Statement to an Information Need Statement via various cognitive activities:

- Listing four research questions (Question 6)
- Visualizing their
  - Topic statement (Questions 4 and 5)
  - The most important research question (Question 8)
- Ranking
  - Concept terms related to their topic (Question 3)
  - The four research questions (Question 7)
  - Concept terms related to the information need (Question 10)

The interview schedule requires the subject to visualize concept terms related to the Topic Statement in two different ways. We then asked the subject which way he or she prefers. This was to gain some notion of the efficacy of circle (for concepts) and line (for relationships between concepts) visualization over other types of visualizations, which we may use in the Associative Index. The third visualization, of the concept terms related to the preferred research question, was not specified to the student; they were free to "map out" the terms as they wished.

The study tests in a very general sense the efficacy of these activities for promoting cognition in the subject, using such efficacy variables as subject preference or by counting and comparing the number of concept terms listed by the subject at various points in the interview schedule. The cognitive activities we asked the subjects to carry out in the interview schedule were always in the same order (for all subjects), which is a limitation of the research instrument.

Results

The purpose of the interview schedule was to narrow the undergraduates' topic statements about their course essay to an information need focus, so as to facilitate user identification of their information need. The schedule does this via various points of cognition, which require the undergraduate to carry out various listing, visualization, ranking, narrowing, and selecting activities. Research questions reported here are derived from expected changes in subject behavior due to these activities.

TABLE 1. Number of terms in topic versus information need statement.

Subject: Course level	N =	Topic Statement: No. of terms derived from . . .	Information need statement: No. of terms derived from . . .
History: 200-level	**20**	**115**	**135**
History: 400-level	(a) 9	63	63
	(b) 10	71	85
History: 400-level Subtotal	**(a) + (b) = 19**	**134**	**148**
Psychology: 400-level	**20**	**114**	**121**
Total	**59**	**363**	**404**

- Research Question 1: Will there be a net reduction in the number of terms derived from the Information Need Statements compared to the number of terms derived from the initial Topic Statements?

Table 1 indicates that from the initial Topic Statements of the 59 students to their final Information Need Statements, the total number of terms increased from 363 to 404.

- Research Question 2: When the terms derived from the Topic Statement are compared to the terms derived from the Information Need Statement, will the majority of terms stay the same in both lists?

Table 2 indicates that the largest category of terms were added terms, with omitted terms and terms that stayed the same being about equal in both term lists.

- Research Question 3: Will there be no change in the students' rank ordering of terms derived from their Topic Statements when they are asked specifically to reconsider all terms in the list and then to rank them in order of importance?

Students were asked to (a) list terms from their Topic Statement, and then (b) rank order the terms from the point

TABLE 2. Same, added, and omitted terms in the information need statement term list compared to original topic statement term list.

	N =	Same	Added	Omitted
History: 200-level	**20**	**43**	**56**	**36**
History: 400-level	(a) 9		33	31
	(b) 10		37	24
History: 400-level Subtotal	**19**	**23**	**70**	**55**
Psychology: 400-level	**20**	**62**	**31**	**28**
Total =	**59**	**128**	**157**	**119**

TABLE 3. Aggregate ranking given student-listed terms derived from original topic statement.

Listed term ranking	1	2	3	4
1st	**34**	10	7	2
2nd	7	**18**	12	6
3rd	3	10	**13**	5
4th	0	4	0	**10**
5th	2	2	2	3
6th	0	1	3	2
7th	1	2	1	2

Note. Rows and columns do not add up to total number of subjects in the study because the table only includes four listed terms.

of view of importance in their paper. Table 3 indicates that most students, when asked to rank the listed terms taken from their Topic Statement, rank the first on the list as first in importance (34 students). However, a sizeable minority of students ranked the second and third terms, and to a lesser extent the fourth listed term, as first in importance (10, 7, and 2 students, respectively).

- Research Question 4.1: When the students are asked to visualize their topic terms in a Circle Line Visualization (CLV), will there be no changes in the number of terms compared to the Original List (OL) of terms?

Table 4.1 indicates that contrary to the prediction in the research question, some terms were added (six terms) or were omitted (five terms) when the students visualized their topic using a Circle Line Visualization (CLV), but very few terms were changed.

- Research Question 4.2: When the students are asked to visualize their topic terms in a Circle Line Visualization (CLV), will none of the students make changes in the terms compared to the Original Visualization (OV)?

Table 4.2 indicates that contrary to the prediction in the research question, there were some students who added (5 students) or omitted (4 students) terms when asked to change the form of their visualization to a Circle Line Visualization (CLV), but most students in the study made no changes (52 of 59 students).

TABLE 4.1. # Students who added or omitted terms (CLV) compared to Original Listing (OL) of terms from topic statement.

	N =	OL	CLV #added	CLV #omitted
History: 200-level	**20**	N/A	**0**	**1**
History: 400-level	9	N/A	0	0
	10	N/A	1	3
History: 400-level Subtotal	**19**	N/A	**1**	**3**
Psychology: 400-level	**20**	N/A	**5**	**1**
Total =	**59**	N/A	**6**	**5**

TABLE 4.2. Number of students who added, omitted, and/or made no changes in terms in the Circle Line Visualization (CLV) compared to their Original Visualization (OV).

	N =	OV	CLV #added	CLV #omitted	CLV #No change
History: 200-level	**20**	N/A	**0**	**1**	**19**
History: 400-level	9	N/A	0	0	9
	10	N/A	1	1	8
History: 400-level Subtotal	**19**	N/A	**1**	**1**	**17**
Psychology: 400-level	**20**	N/A	**4**	**2**	**16**
Total =	**59**	N/A	**5**	**4**	**52**

- Research Question 4.3: When the students are asked to visualize their topic terms in a Circle Line Visualization (CLV), will there be no changes in the relationships between terms compared to the Original Visualization? (Relationships between terms are lines or lines with arrows drawn by the student to connect one term to another.)

Table 4.3 indicates that when all the students' term relationships are considered, almost twice as many relationships were added in Circle Line Visualization as were omitted from Original Visualization.

In a further analysis of Table 4.3, Table 4.3.1 indicates that 44 of the 59 students added new relationships in the Circle-Line Visualization compared to their Original Visualization, and 26 of 59 students omitted relationships (certain students did both). Eight of 59 students made no changes, but 6 of these 8 were grouped in the 3rd- year psychology course. In addition, the propensity to add and omit relationships was less pronounced for the 3rd-year psychology students than for the 3rd-year history students.

- Research Question 5: When asked to visualize their topic as they wish, will most students draw a cognitive map type diagram using circles (of various sizes to indicate importance of each term relative to other terms) and lines (to connect linked terms)?

Table 5 indicates that students more frequently use arrow diagrams as their default diagram style. This means they write terms out and then use arrows to link the written terms.

TABLE 4.3. Total number of added and omitted relationships between terms in Circle Line Visualization (CLV) compared to students' Original Visualization (OV).

	N = 59	OV	CLV #added	CLV #omitted
History: 200-level	**20**	N/A	**56**	**21**
History: 400-level	(a) 9	N/A	34	19
	(b) 10	N/A	36	17
History: 400-level (a) + (b) Subtotal	**19**	N/A	**70**	**36**
Psychology: 400-level	**20**	N/A	**36**	**22**
Total =	**59**	N/A	**162**	**79**

TABLE 4.3.1. Number of students who added and/or omitted and/or made no change in relationships between terms in the Circle Line Visualization (CLV), as compared to their Original Visualization (OV).

	N =	OV	CLV #added	CLV #omitted	CLV #No change
History: 200-level	**20**	N/A	**15**	**7**	**1**
History: 400-level (2 courses)	9	N/A	9	5	0
	10	N/A	8	5	1
History: 400-level Subtotal	**19**	N/A	**17**	**10**	**1**
Psychology: 400-level	**20**	N/A	**12**	**9**	**6**
Total	**59**	N/A	**44**	**26**	**8**

- Research Question 6: Will students select Circle-Line Visualization over all other listing and visualization approaches as the most useful approach for finding information for their essay?

Table 6 shows that students in the study had a preference for the list of terms derived from their Information Need Statement over all other lists and visualization approaches. However, the 1st-year history subjects indicated the four-question visualization (QV) as the preferred approach; while the 3rd-year psychology subjects indicated a preference for the circle-line visualization (CLV) approach.

Discussion

The Results in the previous section are presented as six research questions, each predicting the efficacy of various listing, ranking, visualizing, and narrowing activities. These activities constitute some of the key points of cognition contained in the interview schedule, shown in The Appendix. We will discuss the results, grouping the research questions according to type of cognitive activity.

Research Questions 1 and 2 make predictions about the efficacy of the narrowing focus of the entire interview schedule, from a general Topic Statement to a narrower Information Need Statement. The Research Questions assume (1) that the students' last statement at the end of the interview schedule about the information they need is the student's Information Need Statement, and (2) that an Information Need Statement is a narrower version of the Topic Statement,

TABLE 5. Style of visualization diagram chosen by students themselves for the Original Visualization (OV).

Style of visualization diagram	History: 200-level	History: 400-level	Psychology: 400-level	Total
1. Venn	2 (10%)	1 (4.7%)	1 (5%)	4 (7%)
2. Bubbles				
3. Circles/lines	2 (10%)	3 (14.2%)	5 (25%)	10 (16.5%)
4. Outline				
5. Tree	1 (5%)	3 (14.2%)	2 (10%)	6 (10%)
6. Flowchart	6 (30%)	1 (4.7%)		7 (11.5%)
7. Arrowgram	6 (30%)	6 (28.5%)	7 (35%)	19 (31%)
8. Boxes		2 (9.5%)		2 (2%)
9. Combination	3 (15%)	5 (23.8%)	5 (25%)	13 (22%)

TABLE 6. Subjects' preferred approaches among Original List (OL), Original Visualization (OV), Circle Line Visualization (CLV), four-Question Visualization (QV), and Last List (LL).

Term list or visualization type	History: 200-level	History: 400-level	Psychology: 400-level	Total
OL	3 (16%)	1 (5.5%)	1 (5%)	5 (8%)
OV	2 (10.5%)	4 (22.2%)	3 (15%)	9 (16%)
CLV	3 (16%)	4 (22.2%)	8 (40%)	15 (26.5%)
QV	6 (31.5%)	3 (17%)	2 (10%)	11 (19%)
LL	5 (26%)	6 (33%)	6 (30%)	17 (30%)

involving fewer terms. With regards to assumption (2): The students in the study derived a total of 363 terms from their Topic Statements and 404 terms from their Information Need Statements, an overall increase in terms derived from the Information Need Statement. Also, while there was a solid core of terms that remained the same in the two listings, more terms were added in the Information Need Statement term listing than were left unchanged or were omitted. The study did not test the first assumption (whether or not the student's last statement is indeed the actual Information Need Statement).

If the first assumption is true, the results indicate that the Information Need Statement is a different conception or perspective from the Topic Statement rather than a narrower version of the Topic Statement with the superfluous terms omitted. For the first time, this gives quantitative corroboration to previous research (e.g., Bates, 1989) that the user identifies his or her information need by interacting with information, thinking about information and undergoing an information process, and that an information process involves the transformation of the user's knowledge structure (i.e., perspective) about the information task (i.e., the student's essay) rather than incremental notions of information acquisition (cf. Brookes, 1980; Cole, 1994; 1997; Todd, 1999). (This is an exciting new area of information science research. For a research study measuring the change in the number of student concept terms over a longer period of time (3 months between measurements), see Pennanen and Vakkari (2003). For an example of research that also quantifies change in users' knowledge structure (according to uncertainty fluctuations), but over Kuhlthau's (1993) six-stage ISP model, cf. Wilson, Ford, Ellis, Foster, & Spink, 2002.)

Research Question 3 tests the efficacy of a "ranking" activity, where the students were asked to consider together all the terms from their Topic Statement, to reconsider them, then to systematically rank order their importance to their essay topic. This Research Question, predicting no change in ranking, proved to be largely true, with 34 of 46 students retaining their first listed term as first in importance to their essay topic. The second and third listed terms, however, were subject to a substantial change in rank ordering. For example, after reconsideration, 29 of the 47 students in the study changed their second listed term ranking. For the third listed term, a substantial minority of 1st-year history students and 3rd-year psychology students actually raised the

ranking of this term to either first or second place. This reordering in ranking may indicate that the activity of systematically considering and ranking previously listed topic terms facilitates the student's cognition.

Research Questions 4.1, 4.2, and 4.3 test the efficacy of a "visualizing" activity, where the students were required to visualize their essay topic using circles and lines. In the study, we compared the student's second visualization, where the student was asked to visualize his or her topic using a Circle Line Visualization (CLV):

- To the student's Original Listing (OL) of terms from the Topic Statement
- To the student's Original Visualization (OV) of those same Topic Statement terms
- To the student's relationship links in the OV

Tables 4.1–4.3 indicate that when the students were asked to diagram their Topic Statement terms using a Circle Line Visualization (CLV) there was very little change in the total number of terms added or omitted by all students in the study compared to either the Original List (OL) of terms or the Original Visualization (OV). However, there was a substantial change in relationships when the students' CLVs were compared to their OVs, with 162 new relationships added to the CLV and 79 OV relationships omitted from the CLV. This despite the fact that, as Table 5 indicates, a substantial minority of the students in the study (16.5%) used a CLV for their OV—requiring this group of students to diagram their essay topic a second time would have had less cognitive effect (although just repeating something twice most likely does have some cognitive effect). Table 4.3.1 indicates that 8 of 59 students neither added nor omitted relationships.

Tables 5 and 6 indicate, respectively, the students' choice of visualization approach if left to their own devices (OV), and their preferred approach to finding information for their essay from among all options, either by the straight listing of terms or by visualizing the terms presented to them in the interview schedule. When left to their own devices, students most frequently used arrow diagrams as their default style for visualization. As the preferred approach to finding information, however, they most frequently chose the list of terms derived from their Information Need Statement rather than other listings or any of the visualizations. Typical of the comments for selecting the Information Need Statement term listing (the last cognitive activity the student was required to do and the final element of the interview schedule) was: "As the final element, the process before allowed me to clarify my thoughts and more easily pick out key concepts to search." Previous research has found this same preference for word lists over visualizations by a large percentage of people (e.g., Allen, 1998; 2000; cf. also, Cole, Mandelblatt, & Stevenson, 2002; Rorvig & Lunin, 1999). This could be related to the cognitive style of the person seeking information. For example, Ellis, Ford, and Wood (1992) found that study subjects previously defined as holists made greater use of a cognitive map when using a hypertext system while previously defined serialists preferred a keyword index (cf. also, Ford, Wilson, Foster, Ellis & Spink, 2002). When designing the INIIReye System, the preference for word lists over visualizations by many potential users of the System must be built into the system.

As in all the other tables, the two preference tables (Tables 5 and 6) just discussed show substantial variations in preferences in information approach between the history and psychology students, and the 1st- and 3rd-year students. For example, the 1st-year history students prefer the four-question visualization (QV) while the 3rd-year psychology students prefer the Circle Line Visualization (CLV) approach. This could be because the 1st-year history students found the four-question visualization (QV) activity helpful in narrowing their essay topic, while the 3rd-year psychology students, who preferred the CLV approach, were already familiar with concept maps (which may also use circles/squares and lines e.g., Lanzing, 1997). In Fall 2003 we will use a 1st-year psychology class with an essay requirement to systematically analyze the differences between groups of students, based on a 2×2 cell structure study design.

Conclusion

In this article we set out components of the INIIReye System, an information need identification information retrieval system for domain novice users in the pre-information need identification stage of their ISP. Information need identification for domain novice users is assumed to occur as a result of interacting with the domain topic as it is defined by the subject discipline. We discuss both library and Internet adaptations of cataloging and classification schemes that bring and place together, respectively, similar information objects for the purpose of facilitating the user's identification of his or her information need. The specific problem addressed in the study is establishing an index structure for hypertext and hyperlinking on the Internet, which will serve as a vehicle for the undergraduate user to put together his/her associated thinking about the topic for which information is being sought. The Associative Index first collects the undergraduate's associated thinking on his or her essay topic, via what we call a tacit knowledge collector, and then creates an index to these thoughts, which freezes or snapshots the thinking, making it available for both user and system manipulation. In the present study, only certain parts of the tacit knowledge collector were tested for efficacy in promoting student cognition for information need identification.

The tacit knowledge collector or collection part of the Associative Index facilitates the undergraduate putting together all the thinking associated with his or her essay topic, and then to index it using the Information Need Statement as a starting point for indexing. From there, we hope to establish the structure of the undergraduate's information need via an appropriate essay style, such as the Compare and Contrast Essay style shown on the right-hand side of Figure 1 (Cole, Leide, Beheshti, Large, & Brooks, in press).

Our overall assumption is, after taking undergraduates through the cognitive activities described here, and structuring the disparate data bits of their associated thoughts about their essay topic into an index, that the Associative Index will show them where they have reached in their thinking about the essay topic. With a guiding image in their mind's eye, undergraduates can then effectively query the information source and make relevance judgments, based on their real information need.

This study is a first step to determining if the cognition points—the various visualization, ranking, narrowing, and selecting activities—achieve their purpose, that is, whether they facilitate undergraduates' identification of their information need while they are accessing information using an IR system. This study tested various research questions about the information identification part of the INIIReye System. Adding information accessing and information interaction is the objective of a later phase of the research.

The findings indicate several interesting things about the relative efficacy of the cognitive points or activities in the interview schedule for the promotion of information need-focused cognition. The most interesting is that cognition for information need identification is most facilitated by asking students to articulate relationship links between concept terms when they draw their Circle Line Visualization (CLV) (see Table 4.3). This finding requires further corroboration. If proved true, Figure 1 requires slight modification to emphasize the links between conceptual terms. The most important finding in the study, however, is that it provides the first quantitative evidence in support of the cognitive school's theoretical notion that information is not data-like or incremental in nature, but rather is a process that transforms the information seeker's knowledge structure.

Acknowledgments

The first four authors wish to acknowledge the financial support of the Social Sciences and Humanities Research Council (SSHRC) of Canada, File No. 410-2002-1394. The authors would also like to thank Leanne Bowler for creating the results tables.

References

Allen, B.L. (1998). Visualization and cognitive abilities. In P.A. Cochrane & E.H. Johnson (Eds.), Visualizing subject access for 21st century information resource: 34th clinic on library application for data processing, 1997 (pp. 63–79). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois.

Allen, B.L. (2000). Individual differences and the conundrums of user-centered design: Two experiments. Journal of the American Society for Information Science, 51(6), 508–520.

Bates, M.J. (1986). Subject access in online catalogs: A design model. Journal of the American Society for Information Science, 37(6), 357–376.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5), 407–424.

Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. The Canadian Journal of Information Science, 5, 133–143.

Belkin, N.J., Oddy, R.N., Brooks, H.M. (1982). ASK for information retrieval. Part I: Background and theory. Journal of Documentation, 38(2), 61–71.

Bjorklund, L., Olander, B., & Smith, L.C. (1989). The personal Hypercatalog. ASIS'89. In J. Katzer & G.B. Newby (Eds.), Proceedings of the 52nd ASIS Annual Meeting (Vol. 26, pp. 115–120). Medford, NJ: Learned Information.

Brajnik, G., Mizzaro, S., Tasso, C., & Venuti, F. (2002). Strategic help in user interfaces for information retrieval. Journal of the American Society for Information Science and Technology, 53(5), 343–358.

Brookes, B.C. (1980). The foundations of information science. Part I. Philosophical aspects. Journal of Information Science, 2, 125–133.

Bush, V. (1945). As we may think. Atlantic Monthly, 176(1), 101–108.

Carlyle, A. (1997). Fulfilling the second objective in the online catalog: Schemes for organizing author and work records into usable displays. Library Resources & Technical Services, 41(2), 79–100.

Cole, C. (1994). Operationalizing the notion of information as a subjective construct. Journal of the American Society for Information Science, 45(7), 465–476.

Cole, C. (1997). Calculating the information content of an information process for a domain expert using Shannon's Mathematical Theory of Communication: A preliminary analysis. Information Processing and Management, 33(6), 715–726.

Cole, C., Leide, J.E., Beheshti, J., Large, A., & Brooks, M. (in press). Investigating the ASK hypothesis in a real-life problem situation: A study of history and psychology undergraduates seeking information for a course essay. Journal of the American Society for Information Science and Technology.

Cole, C., Mandelblatt, B., & Stevenson, J. (2002). Visualizing a high recall search strategy output for undergraduates in an exploration stage of researching a term paper. Information Processing & Management, 38, 37–54.

Cutter, C.A. (1876/1904). Rules for a dictionary catalog (3rd ed.), Washington, DC: U.S. Government Printing Office.

Cutter, C.A. (1891–93). Expansive classification: Part 1: The first six classifications. Boston, MA: Cutter.

Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In J. Glazier & R. Powell (Eds.), Qualitative research in information management (pp. 61–83). Englewood, CO: Libraries Unlimited.

Drabenstott, K.M. (2003). Do nondomain experts enlist the strategies of domain experts? Journal of the American Society for Information Science and Technology, 54(9), 836–854.

Donovan, J.M. (1991). Patron expectations about collocation: Measuring the difference between the psychologically real and the really real. Cataloging & Classification Quarterly, 13(2), 23–43.

Ellis, D. (1989). A behavioural approach to information retrieval system design. Journal of Documentation, 45, 171–212.

Ellis, D., Ford, N., & Wood, F. (1992). Hypertext and learning styles. Final report of a project funded by the Learning Technology Unit. Sheffield, UK: Employment Department.

Ford, N., Wilson, T.D., Foster, A., Ellis, D., & Spink, A. (2002). Information seeking and mediated searching. Part 4. Cognitive styles in information seeking. Journal of the American Society for Information Science and Technology, 53(9), 728–235.

Google.com. (2003). Google web directory help. Retrieved September 11, 2003, from http://www.google.com/dirhelp.html#whenuse

Hearst, M.A. (1999). User Interfaces and visualization. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.), Modern information retrieval (pp. 257–323). Reading, MA: Addison-Wesley.

Hendry, D.G., & Harper, D.J. (1997). An informal information-seeking environment. Journal of the American Society for Information Science, 48(11), 1036–1048.

Hert, C.A., Jacob, E.K., & Dawson, P. (2000). A usability assessment of online indexing structures in the networked environment. Journal of the American Society for Information Science, 51, 971–988.

Hjorland, B. (1997). Information seeking and subject representation: An activity-theoretical approach to information science. Westport, CO: Greenwood Press.

Kuhlthau, C.C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42(5), 361–371.

Kuhlthau, C.C. (1993). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex Publishing.

Kuhlthau, C.C. (1999). Accommodating the user's information search process: Challenges for information retrieval system designers. Bulletin of the American Society for Information Science, 25(3), 12–16.
Lanzing, J. (1997). The concept mapping homepage. Retrieved July 24, 2003, from http://users.edte.utwente.nl/lanzing/cm_home.htm
Leide, J., Large, A., Beheshti, J., Brooks, M., & Cole, C. (2003). Visualization schemes for domain novices exploring a topic space: The navigation classification scheme. Information Processing & Management, 39(6), 923–940.
Meadow, C.T., Boyce, B.R., & Kraft, D.H. (2000). Text information retrieval systems. San Diego, CA: Academic Press.
Nelson, T.H. (1972/1991). As we will think. In J.M. Nyce, & P. Kahn (Eds.), From Memex to Hypertext. Vannevar Bush and the Mind's Machine (pp. 245–260). Boston: Academic Press.
Nyce, J.M., & Kahn, P. (1989). Innovation, pragmaticism, and technological continuity: Vannevar Bush's Memex. Journal of the American Society for Information Science, 40(3), 214–220.
The Oxford English Dictionary. (1989). Oxford: Clarendon Press.
Pennanen, M., & Vakkari, P. (2003). Students' conceptual structure, search process, and outcome while preparing a research proposal: A longitudinal case study. Journal of the American Society for Information Science and Technology, 54(8), 759–770.
Ranganathan, S.R. (1962). Elements of library classification. Bombay: Asia Publishing House.
Robertson, S. (2001). Problem solving. Hove: Psychology Press.
Rorvig, M., & Lunin, L.F. (1999). Introduction and overview: Visualization, retrieval and knowledge. Journal of the American Society for Information Science, 50(9), 790–793.
Savage-Knepshield, P.A., & Belkin, N.J. (1999). Interaction in information retrieval: Trends over time. Journal of the American Society for Information Science, 50, 1067–1082.
Saracevic, T., Mokros, H., & Su, L. (1990). Nature of interactions between users and intermediaries in online searching: A qualitative analysis. In D. Henderson (Ed.), Proceedings of the 53rd Annual Meeting of the American Society for Information Science (Vol. 27, pp. 47–54). Medford, NJ: Learned Information.
Sayers, W.C.B. (1959). A manual of classification for librarians and bibliographers. London: Andre Deutsch.
Shera, J.H. (1965). Libraries and the organization of knowledge. London: Crosby Lockwood & Son.
Svenonius, E. (2000). The Intellectual Foundation of Information Organization. Cambridge, MA: MIT Press.
Taylor, A.G. (2000). Wynar's introduction to cataloging and classification. Englewood, CO: Libraries Unlimited.
Taylor, R.S. (1968). Question-negotiation and information seeking in libraries. College & Research Libraries, 29(3), 178–194.
Tebbutt, J. (1999). User evaluation of automatically generated semantic hypertext links in a heavily used procedural manual. Information Processing & Management, 35(1), 1–18.
Todd, R.J. (1999). Back to our beginnings: Information utilization, Bertram Brookes and the fundamental equation of information science. Information Processing & Management, 35, 851–870.
University of Maryland. (1998). What is an alphabetico-classed approach? Office of Information Technology. Workstations at Maryland. Retrieved September 11, 2003, from http://www.wam.umd.edu/~aubrycp/project/Sog250qc.html
Weinberg, B.H. (2002). New course design: Classification schemes and information architecture. Bulletin of the American Society for Information Science and Technology, 28(5), 14–17.
Whitmire, E. (2002). Disciplinary differences and undergraduates' information-seeking behavior. Journal of the American Society for Information Science and Technology, 53(8), 631–638.
Wilson, T.D., Ford, N., Ellis, D., Foster, A., & Spink, A. (2002). Information seeking and mediating searching. Part 2. Uncertainty and its correlates. Journal of the American Society for Information Science and Technology, 53(9), 704–715.

Appendix: Interview Schedule January–April 2003

1st sheet	1.	Could you please describe and write down the topic of your paper?
	2.	What words, search terms, keywords or concepts will you use to look for information to write your paper? Please write out the terms.
	2b.	Include any terms from the topic sentence.
	3.	Could you please label these terms, using initials or abbreviations? Could you rank these terms according to their importance in your paper?
2nd sheet	4.	Could you in some way visualize, draw, map or chart these words, terms or concepts in your paper showing the relationship(s) between them, if any? *Use the labels you have chosen. [When finished]: Please write the search term(s) represented by each label on the diagram. Please verbally explain your diagram.
3rd sheet	5.	Could you now visualize these same terms, using circles and lines this time? The circles should vary in size according to importance of terms. Their distance apart should be representative of how closely related the concepts are. The lines should show inter-relationships amongst the terms. *Please use the labels you chose for each term to label the circles rather than write the term in the circle. [When finished]: Please write the search term words on the diagram & please verbally explain your diagram.
4th sheet	6.	Could you please now write down 4 questions you would like answered by your information source?
	7.	Could you please rank these 4 questions by their importance to your paper?
5th sheet	8.	Concerning the question you just ranked as most important, please write out the terms you would use to look for information for that question. Please label these terms with initials or abbreviations. Could you map out these terms in a new diagram? Please use these labels on your diagram. [When finished]: Please write the search term words on the diagram & please verbally explain your diagram.
6th sheet	9.	Having gone through this process, can you now write down what your paper will be about? Feel free to make changes or integrate new thoughts.
	10.	What words, search terms, keywords or concepts would you use now? Please write them down, and rank these terms according to their importance to your paper.
	11.	Of these approaches [the original list, your initial visualization, the circles and lines diagram, the last diagram, or your final list of terms] which would you find most useful in finding information to write your paper? Please mark the one you prefer and verbally explain why would you prefer it.